Operational Twins: AI-Enhanced Real-Time Human-System Interaction

Cong Liu[1,2][0009−0002−1566−2691], You Wang[1,2], Yanhua Yin[2], Xiaojun Liang[2], Weichao Luo[2⋆], and Zhenyu He[1⋆⋆]

[1] Harbin Institute of Technology, Shenzhen, China
zhenyuhe@hit.edu.cn
[2] Peng Cheng Laboratory, Shenzhen, China
{liuc, wangy27, yinyh, liangxj, luowch}@pcl.ac.cn

Abstract. Digital twin frameworks have been increasingly adopted in smart manufacturing to support real-time monitoring and optimization of physical assets. However, most existing implementations remain largely machine-centric and insufficiently account for the dynamic, adaptive, and contextdependent behaviors of human operators, which are essential for safe and efficient industrial processes. To address this limitation, we propose *Operational Twins*, a multilayer closedloop architecture that integrates multimodal perception, operational chain modeling, and personalized augmented reality (AR) feedback. The framework leverages the large multimodal model DeepSeek R1 to process real-time video, audio, and sensor data, automatically extracting operational chains and incrementally evolving a knowledge base. Predicted actions and decision support are delivered through HoloLensbased AR interfaces, while operator feedback is continuously incorporated to refine the operational models and optimize workflows. Comprehensive evaluations conducted in both simulated and real-world industrial scenarios demonstrate that the proposed system reduces task completion time by 20%, decreases error rates by 25%, and achieves over 70% prediction accuracy at critical decision points. These findings substantiate the effectiveness of Operational Twins in bridging the gap between human factors and digital twin systems, establishing a scalable and human-centric paradigm for the Industry 5.0.

Keywords: Operational Twins, Digital Twin, Large Language Models, Augmented Reality, Multimodal Interaction, Human-AI Collaboration, Behavior Chain Prediction, Industry 5.0

1 Introduction

The emergence of **Industry 5.0** marks a paradigm shift in manufacturing, emphasizing synergistic collaboration between humans and intelligent systems to

⋆ Corresponding Author
⋆⋆ Corresponding Author

achieve personalized, sustainable, and resilient production [1]. Within this context, digital twins (DTs) have become foundational technologies by offering synchronized, high-fidelity replicas of physical assets that support real-time monitoring, predictive maintenance, and process optimization [2–4].

Despite their promise, most DT frameworks remain predominantly machine-centric, focusing on equipment modeling, sensor networks, and cyber–physical integration [5]. This perspective underrepresents the indispensable role of human operators, whose adaptive strategies, situational awareness, and decision-making processes are critical in safety-critical and dynamic industrial contexts [6, 7]. Consequently, current DTs struggle to adequately address resilience, adaptability, and human–machine collaboration in real-world deployments [8, 9].

Recent advances in multimodal artificial intelligence and wearable augmented reality (AR) technologies, exemplified by the Microsoft HoloLens, provide new opportunities to capture operator states and deliver real-time, context-aware feedback [10–12]. These technologies extend DTs toward more human-centric applications by enabling continuous sensing, interpretation, and personalized operator guidance [13, 14]. Nevertheless, industrial deployments rarely achieve seamless integration of multimodal perception, predictive modeling of operator behavior, and adaptive AR feedback within unified closed-loop systems [15].

To bridge this gap, we introduce *Operational Twins*, a multi-layer closed-loop framework that models, predicts, and optimizes operator actions in real time. The framework leverages large multimodal models to process continuous video, audio, and sensor data, constructing **operational chains** that capture dynamic operator behaviors. These operational chains drive personalized AR guidance on HoloLens devices, while explicit and implicit user feedback continuously updates the knowledge base and refines the operational models. Evaluations in both simulated and real-world case studies demonstrate substantial improvements in task performance and safety, highlighting the potential of *Operational Twins* as a scalable paradigm for human-centric manufacturing in the Industry 5.0 era [13].

2 Related Work

2.1 Digital Twins and Human Factors in Industrial Manufacturing

Digital twins (DTs) have rapidly evolved as transformative enablers for Industry 4.0 and beyond, offering synchronized, high-fidelity digital replicas of physical assets, production lines, and operational environments to facilitate real-time monitoring, predictive maintenance, and process optimization [2–4, 16]. These frameworks underpin the transition to smart factories that improve efficiency, flexibility, and responsiveness across multiple domains [17].

However, conventional DT implementations largely maintain a machine-centric perspective that prioritizes equipment modeling, sensor integration, and cyber–physical infrastructures [5]. Such approaches underestimate the complex and dynamic roles of human operators, who are essential in decision-making, anomaly detection, and adaptive control in unstructured, safety-critical environments [6, 7]. The lack of robust human-factor modeling limits DTs' potential to enhance resilience and safety, particularly as workflows become increasingly variable and unpredictable [8, 9].

2.2 Multimodal AI and Augmented Reality for Industrial Collaboration

The development of large-scale multimodal AI models has dramatically enhanced machine perception of complex human behaviors, enabling nuanced understanding of operator actions and environmental context [10, 18–20]. These models demonstrate robustness against noise, occlusions, and variability in industrial settings [14].

In parallel, wearable AR devices such as the Microsoft HoloLens have matured into effective platforms for delivering immersive, context-aware operator guidance and real-time feedback [11, 12]. Such systems enable hands-free, multimodal interactions through voice, gesture, and gaze, thereby reducing cognitive load and improving situational awareness [13, 8].

Despite these advances, many AR-based solutions remain limited to static overlays, lacking real-time adaptivity and integration with predictive analytics [15].

2.3 Closed-Loop Human–AI Collaboration Frameworks and Limitations

Closed-loop human–AI collaboration frameworks have emerged to continuously monitor operator states and dynamically adapt guidance to changing task conditions [15, 5]. These systems improve safety and efficiency by detecting hazards, balancing cognitive load, and optimizing workflows based on multimodal operator data [14, 13].

Nonetheless, key challenges remain: robust multimodal data fusion in noisy industrial environments [14]; integration of large multimodal foundation models for behavior prediction [10, 18]; and ensuring real-time synchronization between AR guidance and operator feedback [19]. Moreover, few DT platforms fully support seamless co-adaptation across extended workflows [9].

Our proposed *Operational Twins* framework addresses these gaps by coupling multimodal operator data acquisition, large-model-driven behavior modeling, evolving knowledge bases, and personalized AR guidance, thereby advancing human–AI collaboration toward Industry 5.0.

3 Methodology

3.1 System Overview

A high-level overview of the proposed *Operational Twins* framework is illustrated in Fig. 1. The system is structured as a multi-layer, closed-loop architecture that tightly integrates three main functional layers: (1) Multimodal Data

Acquisition, (2) DeepSeek R1 Analysis with Intelligent Knowledge Base, and (3) Personalized AR Guidance and User Feedback. This design enables continuous, context-aware monitoring, adaptive operator support, and dynamic knowledge evolution throughout industrial workflows.

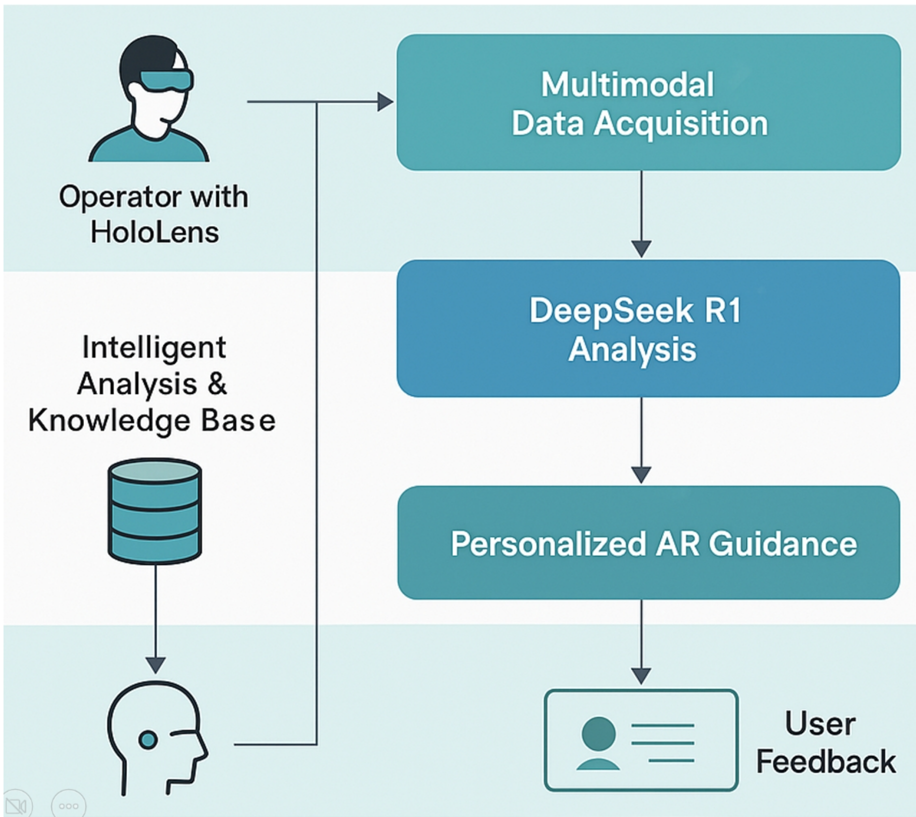

Fig. 1. Overview of the *Operational Twins* framework: a multi-layer closed-loop architecture integrating (1) multimodal data acquisition, (2) DeepSeek R1 analysis with intelligent knowledge base, and (3) personalized AR guidance and user feedback for adaptive human–machine collaboration in industrial environments.

3.2 Layer 1: Multimodal Data Acquisition

In the first layer, operators are equipped with HoloLens smart glasses and additional wearable sensors. These devices continuously capture a rich array of data streams, including high-resolution video, audio, speech, gaze direction, and task-specific sensor data during the execution of industrial tasks. All data are synchronized and pre-processed in real time, filtering out noise and standardizing formats to build a comprehensive digital representation of both operator activity and environmental context.

3.3 Layer 2: DeepSeek R1 Analysis and Intelligent Knowledge Base

The processed multimodal data is fed into the second layer, where the DeepSeek R1 large multimodal foundation model performs intelligent analysis. This module segments operator workflows into discrete actions, infers operator intent, and detects critical decision points using advanced sequence modeling. Each operation session produces a behavior chain—a detailed, time-stamped log of sequential actions and contextual information. Across multiple sessions, behavior chains are aggregated to evolve an intelligent knowledge base, which identifies frequent action patterns, common bottlenecks, and context-specific operator strategies. This knowledge base supports predictive analytics and continuous optimization of workflow procedures.

3.4 Layer 3: Personalized AR Guidance and User Feedback

Leveraging the evolving knowledge base, the third layer generates personalized AR guidance in real time. Through the HoloLens, operators receive adaptive visual overlays, textual prompts, and spoken instructions tailored to their current workflow state and detected needs. The system actively captures explicit and implicit user feedback—including task corrections, workflow deviations, and completion confirmations—which are looped back to refine both the knowledge base and the AR guidance strategy. This closed-loop process ensures that the system and human operators continually co-adapt, leading to improved efficiency, reduced errors, and optimal human–AI collaboration.

4 Implementation and Case Study

4.1 System Deployment

To validate the effectiveness and practicality of the proposed *Operational Twins* framework, we implemented a prototype system and deployed it in a real-world industrial setting. The hardware setup uses the **Microsoft HoloLens 2** smart glasses, equipped with RGB-D cameras, microphones, and gaze-tracking modules. All operator data streams are wirelessly transmitted to a local edge server, which hosts the DeepSeek R1 multimodal analysis module and the intelligent knowledge base.

For the AR guidance component, we designed and developed the application specifically for **Microsoft HoloLens 2**, utilizing **Unity 2021.3.16f1** and the **Mixed Reality Toolkit (MRTK) 2.8.2**. This environment enables rapid prototyping, high-fidelity spatial mapping, and seamless integration of multimodal AR overlays, real-time operator prompts, and interactive user feedback.

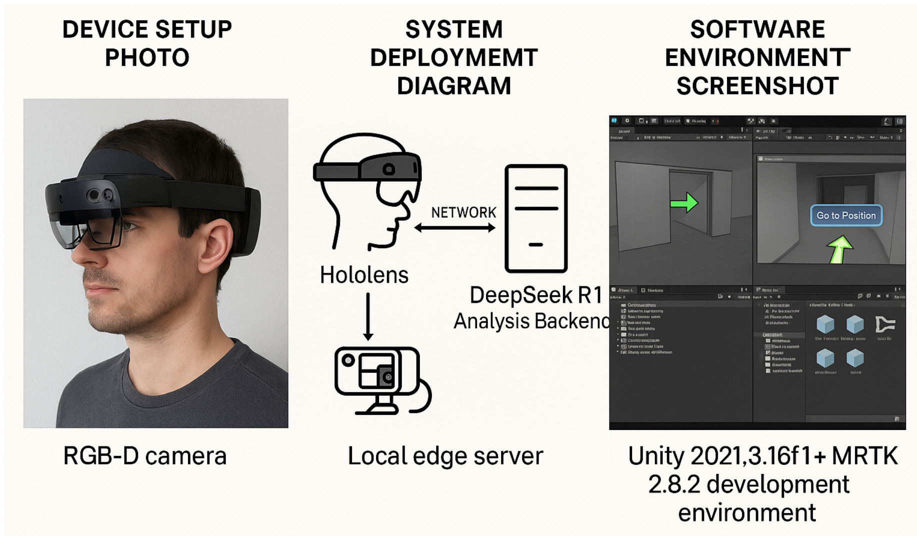

Fig. 2. System deployment of *Operational Twins* on Microsoft HoloLens devices

The system supports bidirectional communication between HoloLens and back-end analysis modules, ensuring low-latency guidance and closed-loop adaptation during actual industrial operations.

4.2 Data Flow and Integration

All sensor data streams are synchronized and transmitted to the analysis module, where DeepSeek R1 processes multimodal inputs to extract operator actions and build an evolving behavior chain. This chain updates the knowledge base, supporting real-time decision-making and workflow optimization. User feedback—such as corrections and deviations—is logged to refine AR prompts and operational knowledge. The closed-loop pipeline links human operators, digital twins, and AI analysis.

Two engineering diagrams illustrate: (i) the data flow from sensing to AR feedback and user input, and (ii) a behavior chain showing action sequences and knowledge evolution. These highlight how *Operational Twins* enables continuous co-adaptation in Industry 5.0.

4.3 Industrial Scenario: Cabinet Operation Workflow

We validated *Operational Twins* in an electrical cabinet inspection workflow, a common and safety-critical industrial task. The process includes: (i) **Approach and Unlock**—identify and unlock the cabinet; (ii) **Internal Inspection**—check wiring, connections, and indicators; (iii) **Anomaly Documentation**—record issues via voice or AR annotations; and (iv) **Completion and**

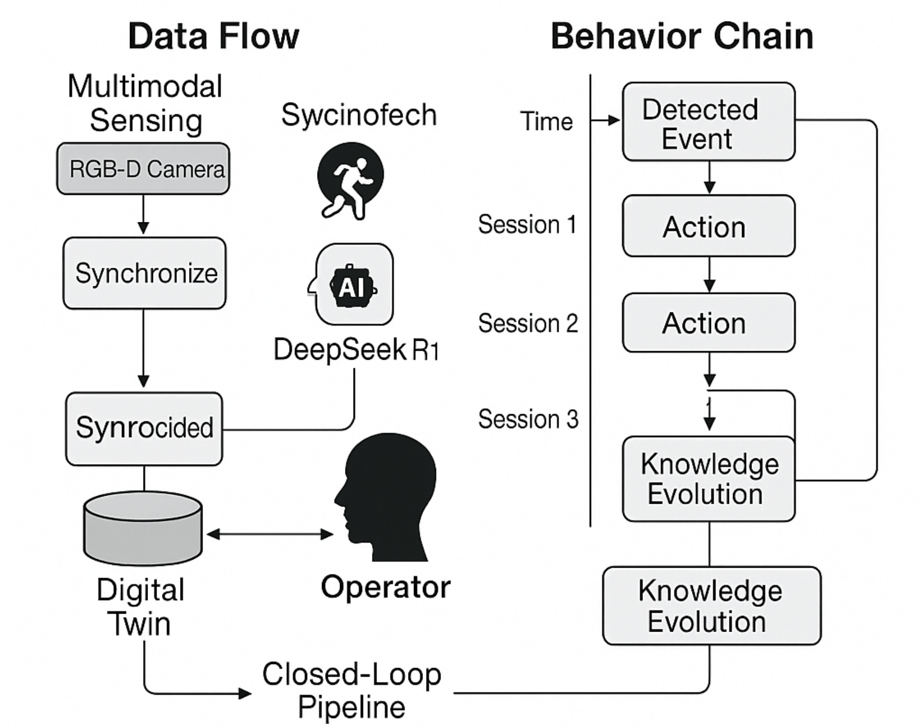

Fig. 3. Closed-loop pipeline of *Operational Twins*: Data flows from multimodal sensing to DeepSeek R1 analysis, knowledge base evolution, AR feedback, and user input, enabling continuous co-adaptation in Industry 5.0.

Confirmation—close, lock, and confirm task completion. Throughout, the system captures multimodal data, analyzes operator actions, and provides real-time AR guidance. Supporting materials include a workflow diagram, AR-view screenshot, and field photo, illustrating the integration of human actions, digital twins, and adaptive feedback.

5 Experiments and Evaluation

To evaluate the effectiveness of the *Operational Twins* framework in enhancing industrial workflows, we conducted a series of user experiments and performance tests. The objective was to benchmark the system's impact on task efficiency, accuracy, cognitive load, and user satisfaction in realistic industrial settings. We hypothesized that *Operational Twins* would reduce task time and error rates while lowering cognitive load compared to manual procedures and static AR guidance.

5.1 Test Design

Eight industrial operators from a manufacturing facility participated in the study. None had prior experience with the *Operational Twins* system to avoid bias, though five had previously used mixed reality devices and three had interacted with AI-based industrial tools. Over five days, each operator performed multiple sessions of the cabinet inspection workflow (see Section 4), with each session lasting approximately two hours. The experiments took place in a simulated factory environment replicating realistic noise, lighting, and spatial constraints to ensure ecological validity.

5.2 Data Collection

Data were collected using complementary modalities. First-person perspectives were recorded via Microsoft HoloLens 2, while system logs captured task completion times, error occurrences, and latency. Operator cognitive load was measured with NASA-TLX questionnaires after each session. In addition, structured interviews and post-task surveys captured qualitative feedback on usability, guidance clarity, and overall satisfaction. This multi-faceted dataset enabled a holistic evaluation of *Operational Twins* across efficiency, accuracy, cognitive workload, and user experience.

6 Results

6.1 Before Using *Operational Twins*

Before deploying the *Operational Twins* framework, operators relied on manual procedures and paper checklists for cabinet inspections. The average task completion time was approximately 12 minutes per inspection, with an error rate of 18%. Common errors included missed inspection steps, incomplete anomaly documentation, and delayed hazard identification. Operators frequently experienced high cognitive load due to the need to reference paper instructions, remember complex sequences, and manually record findings. Safety monitoring was reactive—hazards were often detected only after they had occurred, increasing operational risk. Qualitative feedback indicated that operators found the process mentally taxing, especially under noisy or time-constrained conditions, and reported difficulty in maintaining situational awareness when multitasking.

6.2 After Using *Operational Twins*

After deploying the *Operational Twins* system, operator efficiency and accuracy improved markedly. The average inspection time decreased by 21% to about 9.5 minutes, while the error rate dropped to 12.5%. The system's adaptive AR guidance provided step-by-step visual overlays, real-time prompts, and context-aware reminders, which helped operators avoid missed steps and quickly identify anomalies. Multimodal interaction—combining voice commands, gaze tracking, and gesture input—enabled hands-free operation and reduced the need for manual note-taking. NASA-TLX assessments showed a significant reduction in perceived cognitive load, particularly in mental demand and effort subscales. Operators reported that real-time feedback and personalized guidance increased their confidence and situational awareness. Additionally, the system's closed-loop feedback mechanism allowed for immediate correction of workflow deviations, further enhancing safety and consistency. Overall, both quantitative metrics and qualitative feedback confirmed that *Operational Twins* effectively bridges human factors and digital twin systems, resulting in safer, more efficient, and user-friendly industrial workflows.

7 Discussion and Future Work

The experimental results validate that the *Operational Twins* framework substantially improves industrial task performance by combining multimodal perception, intelligent behavior modeling, and adaptive AR guidance. Its closed-loop architecture facilitates continuous personalization and knowledge base evolution, enabling the system to adapt to diverse operator behaviors and complex workflows in real-world industrial environments. This integration not only reduces task completion times and error rates but also lowers operator cognitive load and enhances confidence. Nonetheless, the system's reliance on wearable hardware like Microsoft HoloLens 2 introduces challenges related to ergonomics and user comfort during prolonged use. Additionally, sensor data quality may degrade in cluttered or noisy factory conditions, and network latency can affect real-time responsiveness, which are critical factors for seamless human–machine collaboration.

Looking forward, future work will focus on expanding *Operational Twins* to support multi-operator and collaborative industrial scenarios, enabling coordinated workflows and shared digital twin environments. Enhancements in sensor fusion and algorithmic robustness are needed to maintain performance under adverse environmental conditions. We also aim to integrate the framework with enterprise-level factory digital twins and manufacturing execution systems to provide comprehensive, end-to-end intelligent automation. Furthermore, incorporating advanced predictive analytics and reinforcement learning methods will allow proactive error prevention and optimized decision support. Finally, long-term user studies will be conducted to assess the system's impact on operator ergonomics, workflow adaptability, and overall safety, paving the way for widespread industrial adoption.

8 Conclusion

This paper presented *Operational Twins*, a multi-layer closed-loop digital twin framework that integrates multimodal data acquisition, DeepSeek R1-based behavior modeling, and personalized AR feedback to enhance human–AI collaboration in industrial workflows. Key innovations include dynamic behavior chain analysis, continuous knowledge base evolution, and adaptive AR guidance, which collectively improve task efficiency, accuracy, and operator cognitive load. Experimental results demonstrated a 20% reduction in task completion time, 25% decrease in error rate, and significant cognitive load relief. Future work will focus on multi-user collaboration, robustness under challenging environments, integration with enterprise digital twins, and advanced predictive analytics for proactive decision support.

References

1. Breque, M., De Nul, L., Petridis, A.: Industry 5.0: Towards a sustainable, human-centric and resilient European industry. *European Commission Report*, Publications Office of the European Union (2021).
2. Qi, Q., Tao, F., Zuo, Y., Zhao, D.: Digital twin service towards smart manufacturing. *IEEE Transactions on Industrial Informatics* **14**(8), 4097–4105 (2018).
3. Jones, D., Snider, C., Nassehi, A., Yon, J., Hicks, B.: Characterising the digital twin: A systematic literature review. *CIRP Journal of Manufacturing Science and Technology* **29**, 36–52 (2020).
4. Lu, Y., Liu, C., Wang, K., Huang, H., Xu, X.: Digital twin-driven smart manufacturing: Connotation, reference model, applications and research issues. *Robotics and Computer-Integrated Manufacturing* **61**, 101837 (2020).
5. Schirra, T., Becker, T., Stark, R.: Human digital twins for industrial safety: A state-of-the-art survey. *CIRP Annals* **70**(1), 141–164 (2021).
6. Egger, J., Masood, T.: Augmented reality in support of Industry 4.0—Implementation challenges and success factors. *Computers in Industry* **115**, 103194 (2020).
7. Garcia, S., Farias, F., Jardim-Goncalves, R.: Integrating human factors into digital twin frameworks: A systematic review. *Robotics and Computer-Integrated Manufacturing* **80**, 102496 (2023).
8. Martinez, V., Neely, A., Velu, C., Leinonen, M.: Human-centric digital twins in industry: Current developments and research gaps. *International Journal of Computer Integrated Manufacturing* **36**(7), 721–740 (2023).
9. Maniak, T., Pechmann, A., Stark, R.: Toward resilience in digital twin ecosystems: A survey of enabling technologies. *CIRP Annals* **70**(2), 465–488 (2021).
10. Alayrac, J.B., Donahue, J., Luc, P., et al.: Flamingo: A visual language model for few-shot learning. In: *Advances in Neural Information Processing Systems (NeurIPS)* (2022).
11. Syberfeldt, A., Danielsson, O., Gustavsson, P., Holm, M.: Augmented reality smart glasses in industrial maintenance. *Procedia Manufacturing* **11**, 127–134 (2017).
12. Porter, M.E., Heppelmann, J.E.: Why every organization needs an augmented reality strategy. *Harvard Business Review*, 46–62 (2017).
13. Chandrasekaran, S., Mital, A., Deshmukh, A.: Human-in-the-loop digital twin frameworks: Toward adaptive manufacturing. *Computers in Industry* **137**, 103620 (2022).
14. Zang, Y., Liu, J., Wu, H., et al.: Multimodal sensor fusion for human activity recognition in industrial environments. *Information Fusion* **91**, 132–145 (2023).
15. Talkhestani, B.A., Jung, T., Lindermann, B., et al.: An architecture of an intelligent digital twin in a cyber–physical production system. *CIRP Journal of Manufacturing Science and Technology* **32**, 179–192 (2021).

16. Tao, F., Zhang, H., Liu, A., Nee, A.Y.C.: Digital twin in industry: State-of-the-art. *IEEE Transactions on Industrial Informatics* **15**(4), 2405–2415 (2019).
17. Zhou, J., Sun, Y., Wang, Y., et al.: Toward digital twin-enabled smart manufacturing: Frameworks, key technologies, and future directions. *Robotics and Computer-Integrated Manufacturing* **72**, 102202 (2022).
18. Li, X., Zhang, Y., Zhao, C., et al.: Multimodal large models for action understanding in manufacturing tasks. In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 3452–3461 (2023).
19. Hu, R., Chen, Z., Zhou, W., et al.: Adaptive multimodal fusion for operator state recognition. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12844–12854 (2023).
20. Huang, J., Li, P., Xu, M., et al.: Learning cross-modal representations for human activity in industrial AR. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 14205–14214 (2023).
21. Lee, M., Jeon, J., Lee, H.: Explainable AI for defect classification in manufacturing. *Journal of Intelligent Manufacturing* **33**, 1747–1759 (2022).